February 2, 2008

# A GENERALIZED ENUMERATION OF LABELED TREES AND REVERSE PRÜFER ALGORITHM

SEUNGHYUN SEO AND HEESUNG SHIN

Abstract. A *leader* of a tree $T$ on $[n]$ is a vertex which has no smaller descendants in $T$. Gessel and Seo showed

$$\sum_{T\in\mathcal{T}_n} u^{(\text{\# of leaders in } T)}\, c^{(\text{degree of 1 in } T)} = u\,P_{n-1}(1,u,cu),$$

which is a generalization of Cayley formula, where $\mathcal{T}_n$ is the set of trees on $[n]$ and

$$P_n(a,b,c) = c\prod_{i=1}^{n-1}(ia+(n-i)b+c).$$

Using a variation of Prüfer code which is called a *RP-code*, we give a simple bijective proof of Gessel and Seo's formula.

## 1. Introduction

A *tree* on $V$ is an acyclic connected graph with vertex set $V$. In 1889, Cayley [1] showed that $|\mathcal{T}_n| = n^{n-2}$ $(n \geq 1)$, called *Cayley formula*, where $\mathcal{T}_n$ is the set of trees on $[n] = \{1,2,\ldots,n\}$. Later, in 1918, Prüfer [2] made the *Prüfer code* which is a bijection between $\mathcal{T}_n$ and $[n]^{n-2}$. Assume that edges are directed toward the vertex 1 and $\mathrm{indeg}_T(i)$ is the indegree of $i$ in $T$. By Prüfer code, we have

$$\sum_{T\in\mathcal{T}_n}\prod_{i\in[n]} x_i^{\mathrm{indeg}_T(i)} = x_1(x_1+\cdots+x_n)^{n-2},$$

which is a generalization of Cayley formula.

A tree is called a *rooted tree* if one vertex has been designated the root. A vertex $v$ in a rooted tree is a *descendant* of $u$ if $u$ lies on the unique path from the root to $v$. By convention, we consider that (unrooted) trees are rooted at the smallest vertex. A vertex $v$ of a rooted tree is called a *leader* if $v$ is minimal among its descendants. Note that 'leader' is the new terminology of 'proper vertex' which was introduced by Seo [3].

Recently, Gessel and Seo [5] showed that

$$\sum_{T\in\mathcal{T}_n} u^{\mathrm{lead}(T)} c^{\deg_T(1)} = u\,P_{n-1}(1,u,cu), \tag{1}$$

2000 *Mathematics Subject Classification.* 05A15.

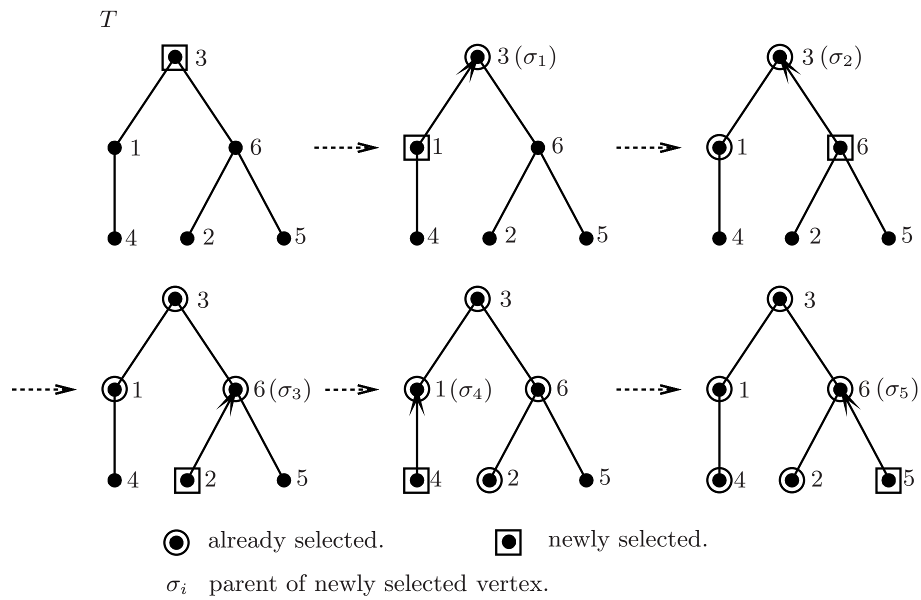


FIGURE 1. Rooted Tree $T$ to RP-code $\varphi(T) = (3, 3, 6, 1, 6)$

where lead$(T)$ is the number of leaders in $T$ and the homogeneous polynomial $P_n(a, b, c)$ is defined by

$$P_n(a, b, c) = c \prod_{i=1}^{n-1} (ia + (n-i)b + c).$$

To prove (1), they used generating functions methods.

In this paper, we prove the equation (1) by giving an algorithm which produces a code with length $n-1$ from a tree with $n$ vertices.

## 2. Reverse Prüfer Algorithm

The *reverse Prüfer code (RP-code)* $\varphi(T) = (\sigma_1, \ldots, \sigma_{n-1})$ of a rooted tree $T$ on $[n]$ is generated by successively selecting the unselected vertex of $T$ having the smallest descendants including itself. If several vertices have the same smallest descendant, we choose the vertex which is the closest to the root. Because the root is selected above all, we assume that the root was already selected. To obtain the code from $T$, we select such a vertex in each step, recording its parent $\sigma_i$, from the tree, until all the vertices are selected. We call this process a *reverse Prüfer algorithm (RP-algorithm)*.

The inverse of $\varphi$ is described as follows: Let $\sigma = (\sigma_1, \ldots, \sigma_{n-1})$ be a sequence of positive integers with $\sigma_i \in [n]$ for all $i$. We can find the tree $T$ whose code is $\sigma$ as building up labeled trees $T_i$ with $i+1$ vertices, except one leaf is unlabeled, by reading the code $\sigma$ forward. Before reading the code, we consider the rooted tree $T_0$ with only one vertex. This root of $T_0$ is

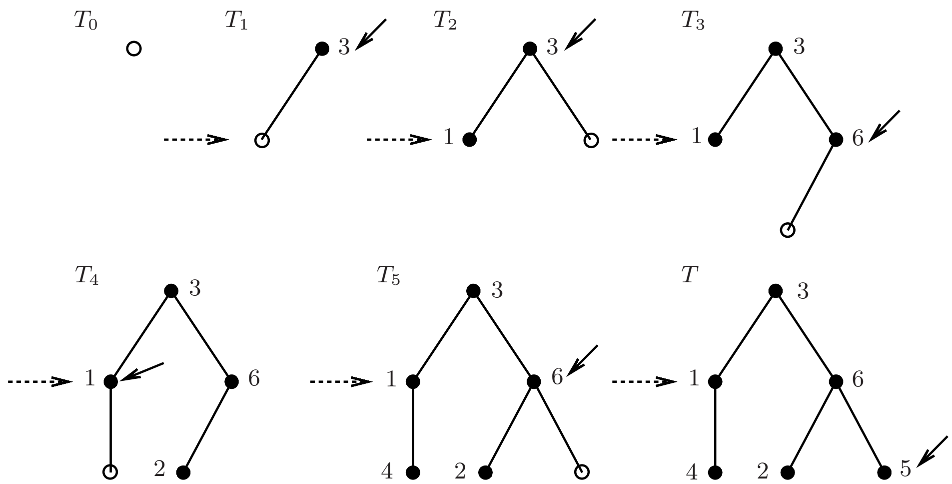


FIGURE 2. RP-code $\sigma = (3, 3, 6, 1, 6)$ to Tree $\varphi^{-1}(\sigma) = T$

unlabeled. Assume that $T_{i-1}$ is the labeled tree which corresponds to initial $i-1$ code $(\sigma_1, \dots, \sigma_{i-1})$ for $i = 1, \dots, n-1$. We make $T_i$ as follows: We label $\sigma_i$ to the unlabeled leaf of $T_{i-1}$. But if $\sigma_i$ is belong to labels of $T_{i-1}$, use the minimum of unused labels instead of $\sigma_i$ as new label number. And then add an unlabeled vertex and an edge between $\sigma_i$ and the just added vertex. After reading the code $\sigma$, we obtain $T_{n-1}$ with $n$ vertices. The unlabeled vertex of $T_{n-1}$ is labeled by the unused label among $[n]$. Then we get the tree $T$. Note that the map $\varphi^{-1}$ was already mentioned in [4, pp. 1–2].

The first coordinate of the RP-code of a rooted tree $T$ is always the label of the root of $T$. In particular, the RP-code of a tree on $[n]$ begins with 1. Cayley formula is reconfirmed by the number of RP-codes.

Figure 3 shows the tree corresponding to the RP-code $(1, 8, 6, 1, 8, 3, 10, 3, 6, 12, 6)$.

## 3. Statistics of Leader in Trees

Now we trace leaders in $T$ during the RP-alogrithm. Let $\sigma = (\sigma_1, \dots, \sigma_n)$ be a RP-code. For each $i = 2, \dots, n$, let $T_{i-1}$ is the tree obtained from subcode $\sigma_1, \dots, \sigma_{i-1}$. Let $l$ be a minimal element in $[n]$ which does not appear in $T_{i-1}$. To construct $T_i$ from $T_{i-1}$ and $\sigma_i$, we should consider the following two cases.

(1) Suppose that $\sigma_i$ appears in $T_{i-1}$. Then the unlabeled vertex $v$ in $T_{i-1}$ is labeled by $l$ in $T_i$. Since the new label $l$ is minimal among unused labels in $T_{i-1}$, the vertex $v$ is a leader in $T$.

(2) Suppose that $\sigma_i$ does not appear in $T_{i-1}$. Then the unlabeled vertex $v$ in $T_{i-1}$ is labeled by $\sigma_i$ in $T_i$.

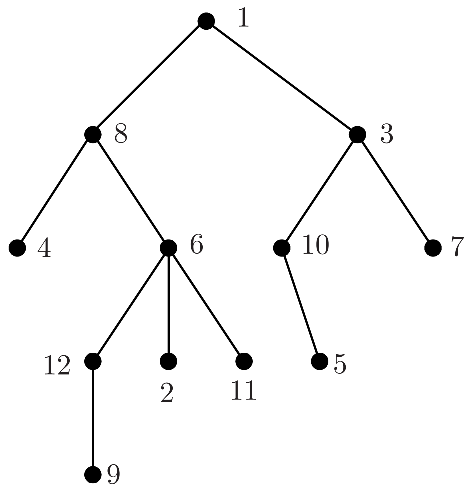


FIGURE 3. Example of the tree with root 1

(a) If $\sigma_i = l$, then the vertex $v$ is leader in $T$ like case (1).
(b) If $\sigma_i \neq l$, then the vertex $v$ has a descendent labeled by $l$. Thus, the vertex $v$ is not leader in $T$.

So there are exactly $i$ choices of $\sigma_i$, case (1) and case (2a), such that the newly labeled vertex $v$ is a leader in $T$. Because the number of $r$'s ($=\sigma_1$) in a RP-code equals to the degree of the root $r$ in $T$, $\deg_T(1)$ is the number of 1 in the RP-code of a tree $T$.

Thus we have the following formula:

$$
\begin{aligned}
\sum_{T \in \mathcal{T}_n} u^{\mathrm{lead}(T)} c^{\deg_T(1)} &= cu && \text{by } \sigma_1 (= 1) \\
&\quad \times ((n-2) + u + cu) && \text{by } \sigma_2 \\
&\quad \times ((n-3) + 2u + cu) && \text{by } \sigma_3 \\
&\qquad \vdots \\
&\quad \times (1 + (n-2)u + cu) && \text{by } \sigma_{n-1} \\
&\quad \times u && \text{by filling the last label} \\
&= cu^2 \prod_{i=2}^{n-1} ((n-i) + (i-1)u + cu). \\
&= uP_{n-1}(1, u, cu).
\end{aligned}
$$

This completes the bijective proof of equation (1).

## 4. Remarks

(1) If $(a_1, \ldots, a_{n-2}, 1)$ is a Prüfer code of $T$ and $\varphi(T) = (1, \sigma_2, \ldots, \sigma_{n-1})$ is a RP-code of $T$, then $a_i = \sigma_{n-i}$ for each $i$. This justifies the terminology *'reverse' Prüfer code.*

(2) With a slight variation of the RP-algorithm, we also find a combinatorial proof of the following formulas for $k$-ary trees and ordered trees.

$$\sum_U u^{\mathrm{lead}(U)} = P_n(k, (k-1)u, u)$$

$$\sum_V u^{\mathrm{lead}(V)} = P_n(1, 2u, u)$$

where $U$ runs all $k$-ary trees and $V$ runs all ordered trees on $[n]$.

## References


1. A. Cayley, *A theorem on trees*, Quart. J. Math. **23**, 376–378, 1889.
2. H. Prüfer, *Neuer Beweis eines Satzes über Permutationen*, Archiv der Math. (3) **27**, 142–144, 1918.
3. S. Seo, *A combinatorial proof of Postnikov's identity and a generalized enumeration of labeled trees*, Electron. J. Combin. **11** no. 2, #N3, 2004.
4. M. Cho, D. Kim, S. Seo and H. Shin, *Colored Prufer Codes for k-Edge Colored Trees*, Electron. J. Combin. **11** no. 1, #N10, 2004.
5. Ira M. Gessel and S. Seo, *A refinement of Cayley's formula for trees*, arXiv: math.CO/0507497, 2005.



Department of Mathematics, Seoul National University, Seoul, 151-742, Republic of Korea

*E-mail address*: `shseo@snu.ac.kr`

Department of Mathematics, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Republic of Korea

*E-mail address*: `H.Shin@kaist.ac.kr`